Multivariate concentration of measure type results using exchangeable pairs and size biasing

Subhankar Ghosh*

January 9, 2010

Abstract

Let $(\mathbf{W}, \mathbf{W}')$ be an exchangeable pair of vectors in $\mathbb{R}^k$. Suppose this pair satisfies

$$E(\mathbf{W}'|\mathbf{W}) = (I_k - \Lambda)\mathbf{W} + \mathbf{R}(\mathbf{W}).$$

If $||\mathbf{W} - \mathbf{W}'||_2 \leq K$ and $\mathbf{R}(\mathbf{W}) = 0$, then concentration of measure results of following form is proved for all $\mathbf{w} \succeq 0$ when the moment generating function of $\mathbf{W}$ is finite.

$$P(\mathbf{W} \succeq \mathbf{w}), P(\mathbf{W} \preceq -\mathbf{w}) \leq \exp\left(-\frac{||\mathbf{w}||_2^2}{2K^2\nu_1}\right),$$

for an explicit constant ν_1, where $\succeq$ stands for coordinate wise $\geq$ ordering.

This result is applied to examples like complete non degenerate U-statistics. Also, we deal with the example of doubly indexed permutation statistics where $\mathbf{R}(\mathbf{W}) \neq 0$ and obtain similar concentration of measure inequalities. Practical examples from doubly indexed permutation statistics include Mann-Whitney-Wilcoxon statistic and random intersection of two graphs. Both these two examples are used in nonparametric statistical testing. We conclude the paper with a multivariate generalization of a recent concentration result due to Ghosh and Goldstein [6] involving bounded size bias couplings and a simple application.

1 Introduction

Stein's method for normal approximation was devised to obtain rates of convergence in central limit theorems. Exchangeable pairs (W, W') satisfying the linearity condition

$$E(W'|W) = (1-\lambda)W \quad \text{for some } \lambda \in (0,1),$$

are often useful for obtaining Kolmogorov distance bounds between the distribution of W and standard normal distribution using Stein's method. The reader is referred to [17] for further details. This condition was generalized in [16] to include a remainder term,

$$E(W'|W) = (1-\lambda)W + R(W), \tag{1}$$

for some measurable function $R(\cdot)$. Using (1), the authors obtained rate of convergence in the central limit theorem for weighted U statistics and antivoter model. Although this condition is quite general, obtaining a usable closed form expression for the remainder term $R(W)$ can be challenging.

*Department of Mathematics, University of Southern California, Los Angeles, CA-90089
2000 *Mathematics Subject Classification*: Primary 60E15; Secondary 60C05,62G10.
Keywords: Large deviations, concentration of measure, exchangeable pairs, Stein's method.

Recently Reinert and Röllin [14] proposed a multivariate formulation of (1). In particular, suppose it is possible to construct an exchangeable multivariate tuple $(\mathbf{W}, \mathbf{W}') \in \mathbb{R}^k \times \mathbb{R}^k$ so that the following relation holds for some matrix Λ and $\mathbf{R} : \mathbb{R}^k \to \mathbb{R}^k$,

$$E(\mathbf{W}'|\mathbf{W}) = (I_k - \Lambda)\mathbf{W} + \mathbf{R}(\mathbf{W}). \tag{2}$$

Under (2), the authors obtain bounds in normal approximation for a rich class of smooth and nonsmooth test functions of $\mathbf{W}$.

Răic [13], Chatterjee [3] and Ghosh and Goldstein [6] obtained concentration of measure type inequalities obtained using tools from Stein's method. Răic used the idea of Cramer transform while Chatterjee used a generalized version of exchangeable pairs. Ghosh and Goldstein [6] obtained concentration results for centered and scaled positive random variables using size biased couplings. In this paper we will obtain some new concentration of measure results under the framework of (2). A general concentration result is contained in Theorem 2.1 for $\mathbf{R}(\mathbf{W}) = 0$, while the case of doubly indexed permutation statistics is also handled later although it does not satisfy this condition.

The paper is organized as follows. In Section 2, we state and prove Theorem 2.1. In Section 3, we apply Theorem 2.1 to obtain concentration of measure results for complete nondegenerate U statistics. In Section 4, we obtain concentration results for doubly indexed permutation statistics which can not be obtained by applying Theorem 2.1. The results for doubly indexed permutation statistics are used to obtain concentration of measure results for two cases of practical importance, the Mann-Whitney-Wilcoxon rank statistic and the random intersection of interpoint distance based graphs, both of which are important in nonparametric hypothesis testing.

2 The main result

In this section and the following, for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$, we define the partial ordering $\succeq$ by

$$\mathbf{a} \succeq \mathbf{b} \Leftrightarrow a_i \geq b_i \quad \text{for } 1 \leq i \leq k.$$

Also, we define the order $\preceq$ by

$$\mathbf{a} \preceq \mathbf{b} \Leftrightarrow \mathbf{b} \succeq \mathbf{a}.$$

The definition for '$\succ$' and '$\prec$' is similar. Also, for any $\boldsymbol{\theta} \in \mathbb{R}^k$, $\boldsymbol{\theta}^t$ stands for transpose. The first theorem of this paper is stated below.

Theorem 2.1. *Suppose $(\mathbf{W}, \mathbf{W}') \in \mathbb{R}^k \times \mathbb{R}^k$ is an exchangeable vector tuple satisfying (2) with $\mathbf{R}(\mathbf{W}) = \mathbf{0}$ that is*

$$E(\mathbf{W}'|\mathbf{W}) = (I_k - \Lambda)\mathbf{W}, \tag{3}$$

for some invertible matrix $\Lambda \in M_k(\mathbb{R})$, the set of $k \times k$ real matrices. Also assume $||\mathbf{W} - \mathbf{W}'||_2 \leq K$ for constant K. If $m(\boldsymbol{\theta}) = E(e^{\boldsymbol{\theta}^t \mathbf{W}}) < \infty$ for all $\boldsymbol{\theta} \in \mathbb{R}^k$, then for any $\mathbf{w} \succeq \mathbf{0}$,

$$P(\mathbf{W} \succeq \mathbf{w}), P(\mathbf{W} \preceq -\mathbf{w}) \leq \exp\left(-\frac{||\mathbf{w}||_2^2}{2K^2\nu_1}\right), \tag{4}$$

where $\nu_1 = 1/\sigma_1(\Lambda)$, with $\sigma_1(\Lambda)$ denoting the smallest singular value of Λ henceforth.

Also the individual coordinate random variables satisfy the following inequalities

$$P(W_i \geq w_i), P(W_i \leq -w_i) \leq \exp\left(-\frac{w_i^2}{2K^2\nu_1}\right) \quad \text{for } i = 1, 2, \ldots, k. \tag{5}$$

Remark 2.1. *If exact value for* ν_1 *is not available, we can use upper bounds on* ν_1 *instead. For example, since*

$$\sigma_1^2(\Lambda) \geq \frac{det(\Lambda^t\Lambda)}{trace(\Lambda^t\Lambda)^{k-1}} = l^2, \tag{6}$$

we obtain $1/l \geq \nu_1$*. Thus we obtain that the right hand side of (4) can be bounded by* $\exp(-(l||\mathbf{w}||_2^2)/2K^2)$.

Before we begin the proof, we note the following inequality which follows by convexity of the exponential function

$$\frac{e^y - e^x}{y - x} = \int_0^1 e^{ty+(1-t)x} dt \leq \int_0^1 (te^y + (1-t)e^x) dt = \frac{e^y + e^x}{2} \quad \text{for all } x \neq y.$$

Hence

$$\frac{|e^{\alpha x} - e^{\alpha y}|}{|x - y|} \leq \frac{|\alpha|(e^{\alpha x} + e^{\alpha y})}{2}. \tag{7}$$

Next we give the proof of Theorem 2.1.

Proof. The gradient vector of $m(\boldsymbol{\theta})$ is given by

$$\nabla m(\boldsymbol{\theta}) = \left(\frac{\partial(m(\boldsymbol{\theta}))}{\partial\theta_i}\right)_{i=1}^k = E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}). \tag{8}$$

Using (3), we obtain,

$$\begin{aligned}
\nabla m(\boldsymbol{\theta}) = E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}) &= E((\mathbf{W} - \mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}}) + E(\mathbf{W}'e^{\boldsymbol{\theta}^t\mathbf{W}}) \\
&= E((\mathbf{W} - \mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}}) + E(E(\mathbf{W}'|\mathbf{W})e^{\boldsymbol{\theta}^t\mathbf{W}}) \\
&= E((\mathbf{W} - \mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}}) + (I_k - \Lambda)E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}).
\end{aligned}$$

Changing sides we obtain

$$\Lambda E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}) = E((\mathbf{W} - \mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}}). \tag{9}$$

Since $(\mathbf{W}, \mathbf{W}')$ is exchangeable, we have

$$E((\mathbf{W} - \mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}}) = E((\mathbf{W}' - \mathbf{W})e^{\boldsymbol{\theta}^t\mathbf{W}'}) = -E((\mathbf{W} - \mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}'}),$$

implying

$$E((\mathbf{W} - \mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}}) = \frac{1}{2}E((\mathbf{W} - \mathbf{W}')(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})). \tag{10}$$

Using (9) and (10), we obtain

$$\Lambda E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}) = \frac{1}{2}E((\mathbf{W} - \mathbf{W}')(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})). \tag{11}$$

Premultiplying both sides by Λ^{-1}, we have

$$E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}) = \frac{1}{2}E(\Lambda^{-1}(\mathbf{W} - \mathbf{W}')(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})).$$

Using (8) and (7), we obtain

$$\begin{aligned}
||\nabla m(\boldsymbol{\theta})||_2 = ||E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}})||_2 &= \frac{1}{2}||E(\Lambda^{-1}(\mathbf{W}-\mathbf{W}')(e^{\boldsymbol{\theta}^t\mathbf{W}}-e^{\boldsymbol{\theta}^t\mathbf{W}'}))||_2 \\
&\leq \frac{1}{2}E(||\Lambda^{-1}(\mathbf{W}-\mathbf{W}')||_2|e^{\boldsymbol{\theta}^t\mathbf{W}}-e^{\boldsymbol{\theta}^t\mathbf{W}'}|) \\
&\leq \frac{1}{2}E(||\Lambda^{-1}||_2||\mathbf{W}-\mathbf{W}'||_2|e^{\boldsymbol{\theta}^t\mathbf{W}}-e^{\boldsymbol{\theta}^t\mathbf{W}'}|) \\
&\leq \frac{1}{4}E\left(||\Lambda^{-1}||_2||\mathbf{W}-\mathbf{W}'||_2|\boldsymbol{\theta}^t(\mathbf{W}-\mathbf{W}')|(e^{\boldsymbol{\theta}^t\mathbf{W}}+e^{\boldsymbol{\theta}^t\mathbf{W}'})\right), \qquad (12)
\end{aligned}$$

where, in the above calculations, for any matrix $A \in M_k(\mathbb{R})$, $||A||_2$ is the spectral norm of A that is

$$||A||_2 = \sup_{\substack{\mathbf{x}\in\mathbb{R}^k \\ ||\mathbf{x}||_2=1}} ||A\mathbf{x}||_2 = \lambda_k^{\frac{1}{2}}(A^tA),$$

where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k$ denote the eigenvalues of A^tA. Denoting $\Lambda^{-t} = (\Lambda^{-1})^t$ and using $(\Lambda^t)^{-1} = (\Lambda^{-1})^t$, we have

$$||\Lambda^{-1}||_2 = \lambda_k^{\frac{1}{2}}(\Lambda^{-t}\Lambda^{-1}) = \lambda_k^{\frac{1}{2}}((\Lambda\Lambda^t)^{-1})) = 1/(\lambda_1^{\frac{1}{2}}(\Lambda\Lambda^t)) = 1/\sigma_1(\Lambda) = \nu_1.$$

Hence, using Cauchy Schwarz inequality, exchangeability of the tuple $(\mathbf{W},\mathbf{W}')$ and $||\mathbf{W}-\mathbf{W}'||_2 \leq K$, (12) yields

$$\begin{aligned}
||\nabla m(\boldsymbol{\theta})||_2 &\leq \frac{||\boldsymbol{\theta}||_2\nu_1}{4}E(||\mathbf{W}-\mathbf{W}'||_2^2(e^{\boldsymbol{\theta}^t\mathbf{W}}+e^{\boldsymbol{\theta}^t\mathbf{W}'})) \\
&\leq \frac{K^2||\boldsymbol{\theta}||_2\nu_1}{4}E(e^{\boldsymbol{\theta}^t\mathbf{W}}+e^{\boldsymbol{\theta}^t\mathbf{W}'}) \\
&\leq \frac{K^2||\boldsymbol{\theta}||_2\nu_1}{2}Ee^{\boldsymbol{\theta}^t\mathbf{W}} = \frac{K^2||\boldsymbol{\theta}||_2\nu_1}{2}m(\boldsymbol{\theta}). \qquad (13)
\end{aligned}$$

Since

$$\nabla(\log(m(\boldsymbol{\theta}))) = \frac{\nabla m(\boldsymbol{\theta})}{m(\boldsymbol{\theta})},$$

we obtain, using (13),

$$||\nabla(\log(m(\boldsymbol{\theta})))||_2 \leq \frac{K^2||\boldsymbol{\theta}||_2\nu_1}{2}. \qquad (14)$$

Hence, using $m(\mathbf{0}) = 1$ and the mean value theorem on $\log(m(\boldsymbol{\theta}))$, we have

$$\log(m(\boldsymbol{\theta})) = \nabla(\log(m(\mathbf{z})))\cdot\boldsymbol{\theta} \leq ||\nabla(\log(m(\mathbf{z})))||_2||\boldsymbol{\theta}||_2, \qquad (15)$$

where $\mathbf{z} \in \mathbb{R}^k$ is a vector in the line segment joining $\mathbf{0}$ to $\boldsymbol{\theta}$. Since (14) holds for any arbitrary $\boldsymbol{\theta} \in \mathbb{R}^k$ and for $\mathbf{z}$ in particular, (15) yields

$$\log(m(\boldsymbol{\theta})) \leq \frac{K^2||\mathbf{z}||_2\nu_1}{2}||\boldsymbol{\theta}||_2 \leq \frac{K^2||\boldsymbol{\theta}||_2^2\nu_1}{2}.$$

Hence

$$m(\boldsymbol{\theta}) \leq \exp\left(\frac{K^2\nu_1||\boldsymbol{\theta}||_2^2}{2}\right). \qquad (16)$$

Hence, for arbitrary $\mathbf{w} \succeq \mathbf{0}$, fixed, for any $\boldsymbol{\theta} \succeq \mathbf{0}$,

$$\begin{aligned} P(\mathbf{W} \succeq \mathbf{w}) &\leq P(\boldsymbol{\theta}^t \mathbf{W} \geq \boldsymbol{\theta}^t \mathbf{w}) \leq e^{-\boldsymbol{\theta}^t \mathbf{w}} m(\boldsymbol{\theta}) & (17) \\ &\leq e^{-\boldsymbol{\theta}^t \mathbf{w}} \exp\left(\frac{K^2 \nu_1 ||\boldsymbol{\theta}||_2^2}{2}\right) = \prod_{i=1}^{k} \exp\left(-\theta_i w_i + \frac{K^2 \nu_1 \theta_i^2}{2}\right). & (18) \end{aligned}$$

We can minimize each term in the product in the right hand side of (18) individually. Using $\theta_i = w_i/(K^2\nu_1)$ in (18), we obtain

$$P(\mathbf{W} \succeq \mathbf{w}) \leq \prod_{i=1}^{k} \exp\left(-\frac{w_i^2}{2K^2\nu_1}\right) = \exp\left(-\frac{||\mathbf{w}||_2^2}{2K^2\nu_1}\right).$$

The other inequality for $P(\mathbf{W} \preceq -\mathbf{w})$ is also derived similarly by considering $\boldsymbol{\theta} \preceq \mathbf{0}$.

Coming to the inequalities for the individual coordinates, take $\boldsymbol{\theta} = (0, \ldots, \theta_i, \ldots, 0)$ that is zero in all coordinates leaving the ith one. Then we obtain

$$P(W_i \geq w_i) \leq e^{-\theta_i w_i} E(e^{\theta_i W_i}) = e^{-\theta_i w_i} m(\boldsymbol{\theta}) \leq e^{-\theta_i w_i} \exp\left(\frac{K^2 \nu_1 \theta_i^2}{2}\right).$$

Letting $\theta_i = w_i/(K^2\nu_1)$ as before yields (5). The left tail bound is similar. □

3 An application from U-statistics

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a vector of i.i.d random variables and $\psi : \mathbb{R}^d \to \mathbb{R}$ be a measurable and symmetric function and $E\psi(X_1, X_2, \ldots, X_d) = 0$. The complete non standardized U-statistics of degree d corresponding to the kernel function ψ is given by

$$U_d(\mathbf{X}) = \sum_{1 \leq j_1 < j_2 < \ldots < j_d \leq n} \psi(X_{j_1}, X_{j_2}, \ldots, X_{j_d}).$$

For $1 \leq k \leq d$, we define following the notations in [15]

$$\psi_k(X_1, X_2, \ldots, X_k) = E\psi(X_1, X_2, \ldots, X_k, X_{k+1}, X_{k+2}, \ldots, X_d | X_1, X_2, \ldots, X_k).$$

If $\mathbf{j} = \{j_1, j_2, \ldots, j_k\} \subset \{1, 2, \ldots, n\}$, we define

$$\psi_k(\mathbf{j}) = \psi_k(X_{j_1}, X_{j_2}, \ldots, X_{j_k})$$

and the corresponding non standardised U statistics is defined by

$$U_k(\mathbf{X}) = \sum_{|\mathbf{j}|=k} \psi_k(\mathbf{j}).$$

Clearly, $U_d(\mathbf{X})$ is the complete nonstandardised U-statistics corresponding to ψ. U-statistics were introduced in [9] and arise naturally in nonparametric statistics. Rinott and Rotar [16] used Stein's method of exchangeable pairs to obtain Kolmogorov distance bounds to normal distribution for weighted U statistics. In [10, 1] concentration of measure results were obtained. While the results in [10] apply to U-statistics of order two only, the results in [1] are very general although applicable to degenerate U-statistics only that is the case when $P(\psi_1(X_1) = 0) = 1$. In the present section, we will obtain concentration of measure results for non degenerate U-statistics and thus will be working with the assumption $P(\psi_1(X_1) = 0) < 1$ henceforth. We will be working with another restriction $||\psi||_\infty \leq b$.

Let us consider the following standardised U statistics for $i = 1, \dots, d$

$$W_i = n^{\frac{1}{2}} \binom{n}{i}^{-1} U_i(\mathbf{X}).$$

It has been shown in [11] that var$W_i \asymp 1$ and furthermore in [15] it was shown that we can embed W_d in a vector $\mathbf{W}$ so that (3) holds. An application of Theorem 2.1 then yields the following result.

Theorem 3.1. *Let $X_1, X_2, \dots, X_n$ be a collection of i.i.d variables. Suppose $\psi : \mathbb{R}^d \to \mathbb{R}$ is a symmetric, measurable function so that $||\psi||_\infty \le b$. Assume $E\psi(X_1, X_2, \dots, X_d) = 0$ and $P(E(\psi(X_1, X_2, \dots, X_d)|X_1) = 0) < 1$. If W_d denotes the U-statistics*

$$W_d = n^{\frac{1}{2}} \binom{n}{d}^{-1} \sum_{|\mathbf{j}|=d} \psi(X_{j_1}, X_{j_2}, \dots, X_{j_d}),$$

then W_d satisfies

$$P(W_d \ge t), P(W_d \le -t) \le \exp\left(-\frac{t^2 \kappa_d^{1/2}}{8b^2 \gamma_d^2}\right),$$

where

$$\gamma_d = \left(\frac{d(d+1)(2d+1)}{6}\right)^{\frac{1}{2}} \quad \text{and} \quad \kappa_d = \frac{(d!)^2 3^{d-1}}{(d(d+1)(2d+1))^{d-1}}.$$

Proof. Let $X'_1, X'_2, \dots, X'_n$ be n independent copies of $X_1, \dots, X_n$. Suppose $\mathbf{X}^i = (X_1, X_2, \dots, X'_i, X_{i+1}, \dots, X_n)$ that is we substitute the ith coordinate with an independent copy of X_i. Define

$$\psi_k^i(\mathbf{j}) = \psi_k(X_{j_1}^i, X_{j_2}^i, \dots, X_{j_k}^i),$$

that is ψ_k applied on the sample with i-th coordinate exchanged. Pick an index I uniformly at random from $1, 2, \dots, n$ and consider the U statistics defined as

$$U'_k = \sum_{|\mathbf{j}|=k} \psi_k^I(\mathbf{j}) \quad \text{and} \quad W'_k = n^{\frac{1}{2}} \binom{n}{k}^{-1} U'_k.$$

It is clear that (W_d, W'_d) is an exchangeable pair, although they do not yield the univariate linearity condition. It has been shown in [15] that with $\mathbf{W} = (W_1, W_2, \dots, W_d)$ and $\mathbf{W}' = (W'_1, W'_2, \dots, W'_d)$, the multivariate Stein condition (3) holds with the lower triangular matrix

$$\Lambda = \frac{1}{n} \begin{pmatrix} 1 & & & & \\ -2 & 2 & & 0 & \\ & -3 & 3 & & \\ & 0 & & \ddots & \ddots \\ & & & -d & d \end{pmatrix}. \tag{19}$$

Clearly $\psi_k(\mathbf{j}) = \psi_k^I(\mathbf{j})$ if $I \notin \mathbf{j}$. Since $||\psi||_\infty \le b$, we therefore obtain,

$$|\psi_k(\mathbf{j}) - \psi_k^I(\mathbf{j})| \le 2b\,\mathbf{1}(I \in \mathbf{j}). \tag{20}$$

Using (20) and $|\{\mathbf{j} : \mathbf{j} \ni I\}| = \binom{n-1}{k-1}$, we have

$$|U_k - U'_k| \le \sum_{\mathbf{j} \ni I} |\psi_k(\mathbf{j}) - \psi_k^I(\mathbf{j})| \le 2b\binom{n-1}{k-1}. \tag{21}$$

Hence we obtain

$$|W_k - W_k'| = n^{\frac{1}{2}} \binom{n}{k}^{-1} |U_k - U_k'| \le n^{\frac{1}{2}} \binom{n}{k}^{-1} 2b \binom{n-1}{k-1} = 2bkn^{-\frac{1}{2}}.$$

The bound above readily yields

$$||\mathbf{W} - \mathbf{W}'||_2 \le 2bn^{-\frac{1}{2}} \left(\frac{d(d+1)(2d+1)}{6} \right)^{\frac{1}{2}} = 2bn^{-\frac{1}{2}} \gamma_d. \tag{22}$$

Using (22) and (19), we can apply Theorem 2.1 with $K = 2b\gamma_d n^{-\frac{1}{2}}$. Next, we have to obtain lower bounds on the singular values of Λ as in (19) following Remark 2.1. It is easy to see

$$\mathrm{trace}(\Lambda^t \Lambda) = \frac{1}{n^2} \left(1 + 2 \sum_{k=2}^{d} k^2 \right) = \frac{1}{n^2} \left(\frac{d(d+1)(2d+1)}{3} - 1 \right) < \frac{d(d+1)(2d+1)}{3n^2}. \tag{23}$$

Also,

$$\mathrm{Det}(\Lambda^t \Lambda) = \mathrm{Det}(\Lambda)^2 = (d!)^2 n^{-2d}. \tag{24}$$

Suppose $0 \le \sigma_1 \le \sigma_2 \le \ldots \le \sigma_d$ denote the k singular values of Λ in order. Using (6), (23) and (24) we obtain

$$\sigma_1^2(\Lambda) \ge \frac{(d!)^2 (3n^2)^{d-1}}{(d(d+1)(2d+1))^{d-1} n^{2d}} = \kappa_d n^{-2}. \tag{25}$$

Hence with $\nu_1 = 1/\sigma_1(\Lambda)$, we obtain $\nu_1 \le \kappa_d^{-1/2} n$. Thus, using Theorem 2.1, we obtain our result. $\square$

4 Doubly indexed permutation statistics

Let $A = \{a_{i,j,k,l} : 1 \le i,j,k,l \le n\}$ be a collection of real numbers such that $a_{i,j,k,l} = 0$ whenever $i = j$ or $k = l$, $a_{i,j,k,l} = a_{i,j,l,k} = a_{j,i,l,k}$ and $\sum_{i \ne j, k \ne l} a_{i,j,k,l} = 0$. We consider the doubly indexed permutation statistic

$$V_1 = \sum_{1 \le s \ne t \le n} a_{s,t,\pi(s),\pi(t)},$$

where π is a permutation chosen uniformly from S_n, the symmetric group of order n. For notational simplicity, we will borrow the notation $a_{i,j,\pi(k),\pi(l)} = a^{\pi}_{i,j,k,l}$ from [14], so that

$$V_1 = \sum_{1 \le s \ne t \le n} a^{\pi}_{s,t,s,t}. \tag{26}$$

These statistics are natural in several nonparametric hypothesis testing problems in statistics. For example, the Mann-Whitney-Wilcoxon signed rank statistic [12] which tests for the equality of distributions of two sets of data or the multivariate graph correlation statistic due to Friedman and Rafsky [4, 5] which tests whether there is significant correlation present among two sets of multivariate vectors. In these cases one is typically interested in obtaining the p-values for V_1 under the null distribution.

In [14, 18], the authors obtained bounds for the error in normal approximation of V_1 using exchangeable pairs and Stein's method. We will be using the exchangeable pair obtained in [14] to prove the following theorem.

Theorem 4.1. *Let $a_{i,j,k,l}$, $1 \le i,j,k,l \le n$ be a collection of real numbers so that $a_{i,j,k,l} = 0$ if $i = j$ or $k = l$, $\sum_{i,j,k,l} a_{i,j,k,l} = 0$ and $a_{i,j,k,l} = a_{i,j,l,k} = a_{j,i,l,k}$ for all i,j,k,l. If $\sup_{i,j,k,l} |a_{i,j,k,l}| \le b$, then with V_1 as in (26), $W_1 = n^{-3/2} V_1$ satisfies the following concentration inequality for all $t > 0$,*

$$P(W_1 \le -t), P(W_1 \ge t) \quad \le \quad \exp\left(-\frac{t^2}{2\phi_{b,n}}\right), \tag{27}$$

where $\phi_{b,n} = (8(2n-1)b^2(6 + 4/n + 1/n^2))/n$.

Proof. We will first construct an exchangeable pair (V_1, V_1') and equivalently (W_1, W_1') where $W_1' = n^{-3/2} V_1'$ and then construct the pair $(\mathbf{W}, \mathbf{W}')$ satisfying (2). Suppose $\tau_{i,j}$ denotes the transposition of i,j that is

$$\tau_{i,j}(k) = k \text{ for all } k \neq i,j \quad \text{and} \quad \tau_{i,j}(i) = j,\ \tau_{i,j}(j) = i.$$

To construct the exchangeable pair, we select two distinct indices I, J uniformly from $\{1, 2, \dots, n\}$. Letting $\pi' = \pi\tau_{I,J}$, we denote

$$V_1' = \sum_{s,t=1}^{n} a^{\pi'}_{s,t,s,t}.$$

Let $\mathbf{V} = (V_1, V_2, V_3)$ and $\mathbf{V}' = (V_1', V_2', V_3')$, where

$$V_i = \sum_{s=1}^{n} a^{(i)}_{s,\pi(s)} \quad \text{and} \quad V_i' = \sum_{s=1}^{n} a^{(i)}_{s,\pi'(s)} \quad \text{for } i = 2, 3, \text{ where}$$
$$a^{(2)}_{s,t} = \frac{1}{n} \sum_{i,j} a_{s,i,t,j} \quad \text{and} \quad a^{(3)}_{s,t} = \frac{1}{n} \sum_{i,j} a_{i,s,j,t} = a^{(2)}_{s,t}.$$

The last equality above implies $V_2 = V_3$ and $V_2' = V_3'$. It has been shown in [14] that the tuple $(\mathbf{V}, \mathbf{V}')$ satisfies

$$E(\mathbf{V}'|\mathbf{V}) = (I_3 - \Lambda)\mathbf{V} + \mathbf{R}', \tag{28}$$

where $\mathbf{R}' = (R_1, 0, 0)$, with

$$R_1 = -\frac{2}{n(n-1)} \sum_{i \neq j} a^{\pi}_{i,j,j,i} = -\frac{2}{n(n-1)} \sum_{i \neq j} a^{\pi}_{i,j,i,j} = -\frac{2}{n(n-1)} V_1, \tag{29}$$

and

$$\Lambda = \frac{2}{n-1} \begin{pmatrix} \frac{2n-1}{n} & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{30}$$

Using (28), we obtain $(\mathbf{W}, \mathbf{W}') = n^{-3/2}(\mathbf{V}, \mathbf{V}')$ satisfies

$$E(\mathbf{W}'|\mathbf{W}) = (I_3 - \Lambda)\mathbf{W} + \mathbf{R}, \tag{31}$$

where Λ is as in (30) and $\mathbf{R} = n^{-3/2}\mathbf{R}'$.

Next we bound $||\mathbf{W}-\mathbf{W}'||_2$ and $\nu_1=\sigma_1^{-1}(\Lambda)$. First we bound $||\mathbf{W}-\mathbf{W}'||_2$. It is easy to verify that

$$\begin{aligned}
V_1'-V_1 &= -\sum_{s=1}^{n}(a^{\pi}_{I,s,I,s}+a^{\pi}_{J,s,J,s}+a^{\pi}_{s,I,s,I}+a^{\pi}_{s,J,s,J}) \\
&\quad +(a^{\pi}_{I,I,I,I}+a^{\pi}_{I,J,I,J}+a^{\pi}_{J,J,J,J}+a^{\pi}_{J,I,J,I}) \\
&\quad +\sum_{s=1}^{n}(a^{\pi}_{I,s,J,s}+a^{\pi}_{J,s,I,s}+a^{\pi}_{s,I,s,J}+a^{\pi}_{s,J,s,I}) \\
&\quad -(a^{\pi}_{I,I,J,J}+a^{\pi}_{I,J,J,I}+a^{\pi}_{J,I,I,J}+a^{\pi}_{J,J,I,I}) \\
&= -\sum_{s=1}^{n}(a^{\pi}_{I,s,I,s}+a^{\pi}_{J,s,J,s}+a^{\pi}_{s,I,s,I}+a^{\pi}_{s,J,s,J}) \\
&\quad +\sum_{s=1}^{n}(a^{\pi}_{I,s,J,s}+a^{\pi}_{J,s,I,s}+a^{\pi}_{s,I,s,J}+a^{\pi}_{s,J,s,I}) \\
&\quad +(a^{\pi}_{I,J,I,J}+a^{\pi}_{J,I,J,I})-(a^{\pi}_{I,J,J,I}+a^{\pi}_{J,I,I,J}), \qquad (32)
\end{aligned}$$

and also

$$V_3'-V_3=V_2'-V_2=-a^{(2)}_{I,\pi(I)}-a^{(2)}_{J,\pi(J)}+a^{(2)}_{I,\pi(J)}+a^{(2)}_{J,\pi(I)}. \qquad (33)$$

The equalities in (32), (33) along with the facts that $|a_{i,j,k,l}|\le b$ and $|a^{(2)}_{s,t}|\le bn$, for all $1\le s,t\le n$ give

$$|V_1'-V_1|\le 8bn+4b \quad \text{and} \quad |V_i'-V_i|\le 4bn \quad \text{for } i=2,3.$$

Thus we obtain

$$||\mathbf{V}-\mathbf{V}'||_2\le((8bn+4b)^2+32b^2n^2)^{\frac{1}{2}}=4b(6n^2+4n+1)^{\frac{1}{2}}.$$

Since $(\mathbf{W},\mathbf{W}')=n^{-3/2}(\mathbf{V},\mathbf{V}')$, we obtain

$$||\mathbf{W}-\mathbf{W}'||_2\le 4bn^{-3/2}(6n^2+4n+1)^{\frac{1}{2}}=4bn^{-1/2}(6+4/n+1/n^2)^{1/2}:=\eta_{b,n}, \text{ say.} \qquad (34)$$

Next, we need to bound ν_1. As in Remark 2.1, we first obtain $\det(\Lambda^t\Lambda)$ and $\mathrm{trace}(\Lambda^t\Lambda)$.

$$\det(\Lambda^t\Lambda)=\det{}^2(\Lambda)=\left(\frac{8(2n-1)}{(n-1)^3n}\right)^2 \quad \text{and} \quad \mathrm{trace}(\Lambda^t\Lambda)=\left(\frac{2}{n-1}\right)^2\left(\frac{(2n-1)^2}{n^2}+4\right)<\frac{32}{(n-1)^2}.$$

Using Remark 2.1, we obtain

$$\sigma_1^2(\Lambda)\ge\frac{\det(\Lambda^t\Lambda)}{\mathrm{trace}^2(\Lambda^t\Lambda)}\ge\left(\frac{2n-1}{4n(n-1)}\right)^2.$$

Hence, with $\nu_1=\sigma_1^{-1}(\Lambda)$ as in Theorem 2.1, we obtain

$$\nu_1\le\frac{4n(n-1)}{2n-1}<2n. \qquad (35)$$

As in the proof of Theorem 2.1, we consider $m(\boldsymbol{\theta})=E(e^{\boldsymbol{\theta}^t\mathbf{W}})$ for $\boldsymbol{\theta}\in\mathbb{R}^3$. The gradient vector is given by

$$\nabla m(\boldsymbol{\theta})=\left(\frac{\partial(m(\boldsymbol{\theta}))}{\partial\boldsymbol{\theta}_i}\right)_{i=1}^{k}=E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}).$$

Using (31), we obtain,

$$\begin{aligned}
\nabla m(\boldsymbol{\theta})=E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}) &= E((\mathbf{W}-\mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}})+E(\mathbf{W}'e^{\boldsymbol{\theta}^t\mathbf{W}}) \\
&= E((\mathbf{W}-\mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}})+E(E(\mathbf{W}'|\mathbf{W})e^{\boldsymbol{\theta}^t\mathbf{W}}) \\
&= E((\mathbf{W}-\mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}})+(I_3-\Lambda)E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}})+E(\mathbf{R}e^{\boldsymbol{\theta}^t\mathbf{W}}).
\end{aligned}$$

Changing sides we obtain

$$\Lambda E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}) \quad = \quad E((\mathbf{W}-\mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}}) + E(\mathbf{R}e^{\boldsymbol{\theta}^t\mathbf{W}}). \tag{36}$$

Since $(\mathbf{W},\mathbf{W}')$ is exchangeable, we have

$$E((\mathbf{W}-\mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}}) = E((\mathbf{W}'-\mathbf{W})e^{\boldsymbol{\theta}^t\mathbf{W}'}) = -E((\mathbf{W}-\mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}'}),$$

implying

$$E((\mathbf{W}-\mathbf{W}')e^{\boldsymbol{\theta}^t\mathbf{W}}) = \frac{1}{2}E((\mathbf{W}-\mathbf{W}')(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})). \tag{37}$$

Using (36) and (37), we obtain

$$\Lambda E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}) \quad = \quad \frac{1}{2}E((\mathbf{W}-\mathbf{W}')(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})) + E(\mathbf{R}e^{\boldsymbol{\theta}^t\mathbf{W}}). \tag{38}$$

Premultiplying both sides by Λ^{-1}, we have

$$E(\mathbf{W}e^{\boldsymbol{\theta}^t\mathbf{W}}) = \frac{1}{2}E(\Lambda^{-1}(\mathbf{W}-\mathbf{W}')(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})) + E(\Lambda^{-1}\mathbf{R}e^{\boldsymbol{\theta}^t\mathbf{W}}). \tag{39}$$

Equating the first coordinates of the vectors on the two sides of (39), we obtain

$$E(W_1e^{\boldsymbol{\theta}^t\mathbf{W}}) = \frac{1}{2}E([\Lambda^{-1}(\mathbf{W}-\mathbf{W}')]_1(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})) + E([\Lambda^{-1}\mathbf{R}]_1e^{\boldsymbol{\theta}^t\mathbf{W}}), \tag{40}$$

where for a vector $\mathbf{X}$, $[\mathbf{X}]_1 := X_1$ or the first coordinate. Since,

$$\mathbf{R} = -\frac{2}{n(n-1)}(W_1,0,0)^t \quad \text{and} \quad \Lambda^{-1}_{1,1} = \frac{n(n-1)}{2(2n-1)},$$

we obtain,

$$[\Lambda^{-1}\mathbf{R}]_1 = -\Lambda^{-1}_{1,1} \times \frac{2W_1}{n(n-1)} = -\frac{W_1}{2n-1}.$$

Thus, (40) now yields,

$$E(W_1e^{\boldsymbol{\theta}^t\mathbf{W}}) = \frac{1}{2}E([\Lambda^{-1}(\mathbf{W}-\mathbf{W}')]_1(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})) - \frac{1}{2n-1}E(W_1e^{\boldsymbol{\theta}^t\mathbf{W}}).$$

Changing sides, we obtain

$$\frac{2n}{2n-1}E(W_1e^{\boldsymbol{\theta}^t\mathbf{W}}) = \frac{1}{2}E\left([\Lambda^{-1}(\mathbf{W}-\mathbf{W}')]_1(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})\right). \tag{41}$$

As before, note that $||\Lambda^{-1}||_2 = \nu_1$. Taking absolute values on both sides of (41) and using (34) and Jensen's inequality, we obtain

$$\begin{aligned}
|E(W_1e^{\boldsymbol{\theta}^t\mathbf{W}})| \quad &= \quad \frac{2n-1}{4n}\left|E\left([\Lambda^{-1}(\mathbf{W}-\mathbf{W}')]_1(e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'})\right)\right| \\
&\leq \quad \frac{2n-1}{4n}E\left(||\Lambda^{-1}(\mathbf{W}-\mathbf{W}')||_2\left|e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'}\right|\right) \\
&\leq \quad \frac{2n-1}{4n}E\left(||\Lambda^{-1}||_2||\mathbf{W}-\mathbf{W}'||_2\left|e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'}\right|\right) \\
&\leq \quad \frac{(2n-1)\eta_{b,n}\nu_1}{4n}E\left|e^{\boldsymbol{\theta}^t\mathbf{W}} - e^{\boldsymbol{\theta}^t\mathbf{W}'}\right|.
\end{aligned}$$

Taking $\boldsymbol{\theta} = (\theta_1, 0, 0)^t$ and using (7), we obtain

$$\begin{aligned}|E(W_1 e^{\theta_1 W_1})| &\leq \frac{(2n-1)\eta_{b,n}\nu_1}{4n} E\left|e^{\theta_1 W_1} - e^{\theta_1 W_1'}\right| \leq \frac{(2n-1)\eta_{b,n}|W_1 - W_1'||\theta_1|\nu_1}{4n} \frac{E(e^{\theta_1 W_1}) + E(e^{\theta_1 W_1'})}{2} \\ &\leq \frac{(2n-1)\eta_{b,n}^2|\theta_1|\nu_1}{4n} \frac{E(e^{\theta_1 W_1}) + E(e^{\theta_1 W_1'})}{2} = \frac{(2n-1)\eta_{b,n}^2|\theta_1|\nu_1}{4n} E(e^{\theta_1 W_1}). \qquad (42)\end{aligned}$$

Using (42) and (34), we obtain

$$|E(W_1 e^{\theta_1 W_1})| \leq \frac{4(2n-1)b^2(6+4/n+1/n^2)|\theta_1|\nu_1}{n^2} E(e^{\theta_1 W_1}).$$

The bound from (35) yields

$$|E(W_1 e^{\theta_1 W_1})| \leq \frac{8(2n-1)b^2(6+4/n+1/n^2)|\theta_1|}{n} E(e^{\theta_1 W_1}).$$

Hence, with $m_1(\theta_1) = E(e^{\theta_1 W_1})$, we obtain

$$|m_1'(\theta_1)| = |E(W_1 e^{\theta_1 W_1})| \leq \frac{8(2n-1)b^2(6+4/n+1/n^2)|\theta_1|}{n} m_1(\theta_1) = \phi_{b,n}|\theta_1|m_1(\theta_1). \qquad (43)$$

It is easy to see that

$$E(V_1) = \frac{1}{n(n-1)} \sum_{s \neq t, u \neq v} a_{s,t,u,v} = 0,$$

implying $m_1'(0) = E(W_1) = 0$ as well. Since $m_1(\theta_1)$ is a convex function, we therefore have $m_1'(\theta_1) \geq 0$ for $\theta_1 \geq 0$ and $m_1'(\theta_1) \leq 0$, for $\theta_1 \leq 0$.

Using (43), we therefore have for $\theta_1 \geq 0$,

$$m_1'(\theta_1) \leq \phi_{b,n}\theta_1 m_1(\theta_1),$$

which on integration, yields

$$\log(m_1(\theta_1)) \leq \frac{\phi_{b,n}\theta_1^2}{2} \quad \text{for } \theta_1 \geq 0.$$

Similar argument holds for $\theta_1 < 0$ as well, yielding

$$m_1(\theta_1) \leq \exp\left(\frac{\phi_{b,n}\theta_1^2}{2}\right) \quad \text{for all } \theta_1.$$

Using Markov's inequality, we have

$$P(W_1 \geq t) \leq e^{-\theta_1 t} m_1(\theta_1) \leq \exp\left(-\theta_1 t + \frac{\phi_{b,n}\theta_1^2}{2}\right) \quad \text{for all } \theta_1 \geq 0.$$

Using $\theta_1 = t/\phi_{b,n}$, we obtain

$$P(W_1 \geq t) \leq \exp\left(-\frac{t^2}{2\phi_{b,n}}\right).$$

The bound for $P(W_1 \leq -t)$ is similar. □

Next we discuss two applications of Theorem 4.1 to distribution free hypothesis testing. The first one is Mann-Whitney-Wilcoxon signed rank statistic, while the second one is the generalised multivariate correlation measure due to Friedman and Rafsky.

4.1 Applications to Mann-Whitney-Wilcoxon statistic

Let $x_1, x_2, \ldots, x_{n_1}$ and $y_1, y_2, \ldots, y_{n_2}$, $n_1 + n_2 = n$ be independent univariate samples from unknown continuous distributions F_X and F_Y respectively. One is interested in testing the hypothesis

$$H_0 : F_X = F_Y \quad \text{vs.} \quad H_1 : F_X \neq F_Y.$$

The MWW test statistic is defined as

$$V_{MWW} = |\{(i,j) : x_i < y_j\}|. \tag{44}$$

We reject H_0 if V_{MWW} is too large or too small, see [12]. The rate of convergence to normality for V_{MWW} was considered in [18] and [14]. Let $\mathbf{z} = (x_1, x_2, \ldots, x_{n_1}, y_1, y_2, \ldots, y_{n_2})$ and $\pi(i)$ denote the rank of z_i. Under H_0, π is clearly a uniform random permutation. For $1 \leq i, j, k, l \leq n$, define

$$a_{i,j,k,l} = \begin{cases} +\frac{1}{2} & \text{if } 1 \leq i \leq n_1,\ n_1 + 1 \leq j \leq n \text{ and } 1 \leq k < l \leq n \\ -\frac{1}{2} & \text{if } 1 \leq i \leq n_1,\ n_1 + 1 \leq j \leq n \text{ and } 1 \leq l < k \leq n \\ 0 & \text{otherwise.} \end{cases} \tag{45}$$

Since

$$\begin{aligned} V_1 = \sum_{s \neq t} a^{\pi}_{s,t,s,t} &= \sum_{1 \leq s \leq n_1, n_1+1 \leq t \leq n} \frac{1}{2}(\mathbf{1}(x_s < y_{t-n_1}) - \mathbf{1}(x_s > y_{t-n_1})) \\ &= \frac{1}{2} V_{MWW} - \frac{1}{2}(n_1 n_2 - V_{MWW}) \\ &= V_{MWW} - \frac{n_1 n_2}{2}, \end{aligned}$$

and $\sum_{i,j,k,l} a_{i,j,k,l} = 0$, we obtain that V_1 is V_{MWW} mean centered and hence instead of evaluating the p values of V_{MWW} under H_0, we might as well obtain the same for V_1. Since $a_{i,j,k,l}$ in (45) satisfies the hypothesis of Theorem 4.1, we can apply Theorem 4.1, to bound the p values of V_1. In particular, using $b = 1/2$ in Theorem 4.1, we obtain the following proposition.

Proposition 4.1. *Let $x_1, x_2, \ldots, x_{n_1}$ and $y_1, y_2, \ldots, y_{n_2}$, $n_1 + n_2 = n$ be independent univariate samples from unknown continuous distributions F_X and F_Y. Let $a_{i,j,k,l}$ be defined as in (45). If π is a permutation chosen uniformly at random and*

$$V_1 = \sum_{s \neq t} a^{\pi}_{s,t,s,t}.$$

Then $W_1 = n^{-3/2} V_1$ satisfies the following inequality for all $t > 0$

$$P(W_1 \geq t), P(W_1 \leq -t) \leq \exp\left(-\frac{t^2 n}{4(2n-1)(6 + 4/n + 1/n^2)} \right),$$

4.2 Random intersection of interpoint distance based graphs

In [4] and [5], notion of association measures like Kendall's τ were extended to multivariate observations using interpoint distance based graphs. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be n i.i.d vector tuples. We are interested in examining the strength of association between X and Y. This is achieved by constructing k minimal spanning trees or k nearest neighbour spanning subgraphs G_1 and G_2 out of the X and Y datapoints respectively. If E_i denotes the edge set of G_i for $i = 1, 2$, then the statistic of interest is

$$\Gamma_1 = \sum_{1 \leq i,j \leq n} \mathbf{1}((i,j) \in E_1) \mathbf{1}(i, j \in E_2).$$

Clearly, a large value of Γ_1 indicates presence of significant association between X and Y. For notational simplicity, let $a_{i,j,k,l} = c_{i,j}d_{k,l}$, where $c_{i,j} = \mathbf{1}((i,j) \in E_1)$ and $d_{k,l} = \mathbf{1}((k,l) \in E_2)$. We need to compare the observed value of Γ_1 with the baseline p value of V_1 where

$$V_1 = \sum_{s \neq t} a^{\pi}_{s,t,s,t} = \sum_{s \neq t} \mathbf{1}((s,t) \in G_1)\mathbf{1}((\pi(s),\pi(t)) \in G_2), \tag{46}$$

where π is a permutation chosen uniformly at random from S_n. Clearly

$$\mu = E(V_1) = \frac{1}{n(n-1)} \sum_{i \neq j, k \neq l} a_{i,j,k,l} = \frac{4|E_1||E_2|}{n(n-1)}.$$

Hence, if we consider

$$\widehat{a}_{i,j,k,l} = \begin{cases} a_{i,j,k,l} - \frac{4|E_1||E_2|}{n^2(n-1)^2} & \text{if } i \neq j \text{ and } k \neq l \\ 0 & \text{otherwise.} \end{cases}$$

then $\sum_{i,j,k,l} \widehat{a}_{i,j,k,l} = 0$ and the array $\widehat{a}_{i,j,k,l}$ $1 \leq i,j,k,l \leq n$ satisfies the conditions in Theorem 4.1. Since $|E_1|, |E_2| \leq n(n-1)/2$, the number of edges in the complete graph on n vertices, we obtain

$$|\widehat{a}_{i,j,k,l}| \leq a_{i,j,k,l} + \frac{4|E_1||E_2|}{n^2(n-1)^2} \leq 2.$$

Hence applying Theorem 4.1 with $b = 2$, we obtain the following proposition.

Proposition 4.2. *Let $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ be two interpoint distance based graphs derived from n data points $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ respectively. Let π be a permutation chosen uniformly at random from S_n. Then V_1, as defined in (46) satisfies the following concentration inequality*

$$P\left(n^{-3/2}\left(V_1 - 4\frac{|E_1||E_2|}{n(n-1)}\right) \geq t\right), P\left(n^{-3/2}\left(V_1 - 4\frac{|E_1||E_2|}{n(n-1)}\right) \leq -t\right) \leq \exp\left(-\frac{nt^2}{64(2n-1)(6+4/n+1/n^2)}\right).$$

5 Size biasing and multivariate concentration inequalities

Let $\mathbf{W} = (W_1, W_2, \ldots, W_k) \in \mathbb{R}^k$, be a random vector with nonnegative coordinate variables. In [6], concentration of measure inequalities were obtained for positive random variable W with positive mean μ and nonzero variance σ^2 under a boundedness condition on the coupling (W, W^s), where W^s denotes the size bias transformation of W, that is, it satisfies the identity

$$E(Wf(W)) = \mu E(f(W^s)) \quad \text{for all functions } f \text{ so that } E(Wf(W)) \text{ is defined.}$$

In this section, we will derive a multivariate analogue of the same result. For $\mathbf{W}$ in consideration, assume $\mu_i > 0$ for all $i = 1, 2, \ldots, k$. The $\mathbf{W}$ size biased variate in direction i denoted by $\mathbf{W}^i$ is defined as the random variable having distribution dF^i with

$$dF^i(x_1, x_2, \ldots, x_n) = \frac{x_i}{\mu_i} dF(x_1, x_2, \ldots, x_n),$$

where $\mathbf{W} \sim dF$. The random variable $\mathbf{W}^i$ thus defined satisfies

$$E(W_i f(\mathbf{W})) = \mu_i E(f(\mathbf{W}^i)),$$

for all functions f where the above expectations are finite. In particular

$$E(W_i e^{\boldsymbol{\theta}^t \mathbf{W}}) = \mu_i E(e^{\boldsymbol{\theta}^t \mathbf{W}^i}). \tag{47}$$

For notational purposes let us define for any two vectors $\boldsymbol{\theta}, \boldsymbol{\phi} \in \mathbb{R}^k$

$$\frac{\boldsymbol{\theta}}{\boldsymbol{\phi}} = \left(\frac{\theta_1}{\phi_1}, \frac{\theta_2}{\phi_2}, \ldots, \frac{\theta_k}{\phi_k}\right).$$

Theorem 5.1. *Suppose* $\mathbf{W} = (W_1, W_2, \ldots, W_k)$ *is a random vector with nonnegative coordinate variables, with* $\boldsymbol{\mu}, \boldsymbol{\sigma} \succ \mathbf{0}$*. Suppose* $||\mathbf{W} - \mathbf{W}^i||_2 \leq K$ *for some constant* K *for all* $i = 1, 2, \ldots, k$*. If* $\sigma_{(1)} = \min_{i=1,2,\ldots,k} \sigma_i$*, then for any* $\mathbf{t} \succeq \mathbf{0}$*, we have*

$$P\left(\frac{\mathbf{W} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \succeq \mathbf{t}\right) \leq \exp\left(-\frac{||\mathbf{t}||_2^2}{2(K_1 + K_2||\mathbf{t}||_2)}\right),$$

where

$$K_1 = \frac{2K}{\sigma_{(1)}}||\frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}||_2 \quad \textit{and} \quad K_2 = \frac{K}{2\sigma_{(1)}}.$$

Proof. Using (7), we obtain for any i,

$$\begin{aligned} E(e^{\boldsymbol{\theta}^t \mathbf{W}^i}) - E(e^{\boldsymbol{\theta}^t \mathbf{W}}) \leq |E(e^{\boldsymbol{\theta}^t \mathbf{W}^i}) - E(e^{\boldsymbol{\theta}^t \mathbf{W}})| &\leq E\left(\frac{|\boldsymbol{\theta}^t(\mathbf{W}^i - \mathbf{W})|(e^{\boldsymbol{\theta}^t \mathbf{W}^i} + e^{\boldsymbol{\theta}^t \mathbf{W}})}{2}\right) \\ &\leq E\left(\frac{||\boldsymbol{\theta}||_2||\mathbf{W}^i - \mathbf{W}||_2(e^{\boldsymbol{\theta}^t \mathbf{W}^i} + e^{\boldsymbol{\theta}^t \mathbf{W}})}{2}\right) \\ &\leq \frac{K||\boldsymbol{\theta}||_2 E(e^{\boldsymbol{\theta}^t \mathbf{W}^i} + e^{\boldsymbol{\theta}^t \mathbf{W}})}{2}. \end{aligned}$$

Changing sides, we obtain for $||\boldsymbol{\theta}||_2 < 2/K$,

$$E(e^{\boldsymbol{\theta}^t \mathbf{W}^i}) \leq \frac{1 + \frac{K||\boldsymbol{\theta}||_2}{2}}{1 - \frac{K||\boldsymbol{\theta}||_2}{2}} E(e^{\boldsymbol{\theta}^t \mathbf{W}}).$$

Hence from (47), we obtain, for $||\boldsymbol{\theta}||_2 < 2/K$

$$\frac{\partial m(\boldsymbol{\theta})}{\partial \theta_i} = E(W_i e^{\boldsymbol{\theta}^t \mathbf{W}}) = \mu_i E(e^{\boldsymbol{\theta}^t \mathbf{W}^i}) \leq \mu_i \frac{1 + \frac{K||\boldsymbol{\theta}||_2}{2}}{1 - \frac{K||\boldsymbol{\theta}||_2}{2}} E(e^{\boldsymbol{\theta}^t \mathbf{W}}) = \mu_i \frac{2 + K||\boldsymbol{\theta}||_2}{2 - K||\boldsymbol{\theta}||_2} m(\boldsymbol{\theta}). \tag{48}$$

Denoting $M(\boldsymbol{\theta}) = E\left(\exp\left(\boldsymbol{\theta}^t \cdot ((\mathbf{W} - \boldsymbol{\mu})/\boldsymbol{\sigma})\right)\right)$, we obtain

$$M(\boldsymbol{\theta}) = m\left(\frac{\boldsymbol{\theta}}{\boldsymbol{\sigma}}\right) e^{-\boldsymbol{\theta}^t \cdot \boldsymbol{\mu}/\boldsymbol{\sigma}}. \tag{49}$$

Hence denoting

$$\partial_i m(\boldsymbol{\beta}) = \left.\frac{\partial m(\boldsymbol{\theta})}{\partial \theta_i}\right|_{\boldsymbol{\theta} = \boldsymbol{\beta}},$$

for $\boldsymbol{\beta} \in \mathbb{R}^k$ and using (49) and (48), we obtain, for $||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2 < 2/K$,

$$\begin{aligned} \frac{\partial M(\boldsymbol{\theta})}{\partial \theta_i} &= \frac{1}{\sigma_i} \partial_i m\left(\frac{\boldsymbol{\theta}}{\boldsymbol{\sigma}}\right) \exp\left(-\boldsymbol{\theta}^t \frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right) - \frac{\mu_i}{\sigma_i} m\left(\frac{\boldsymbol{\theta}}{\boldsymbol{\sigma}}\right) \exp\left(-\boldsymbol{\theta}^t \frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right) \\ &\leq \frac{\mu_i}{\sigma_i} \frac{2 + K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2}{2 - K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2} m\left(\frac{\boldsymbol{\theta}}{\boldsymbol{\sigma}}\right) \exp\left(-\boldsymbol{\theta}^t \frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right) - \frac{\mu_i}{\sigma_i} m\left(\frac{\boldsymbol{\theta}}{\boldsymbol{\sigma}}\right) \exp\left(-\boldsymbol{\theta}^t \frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right) \\ &= \frac{\mu_i}{\sigma_i} M(\boldsymbol{\theta}) \left(\frac{2 + K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2}{2 - K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2} - 1\right) = \frac{\mu_i}{\sigma_i} M(\boldsymbol{\theta}) \frac{2K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2}{2 - K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2}. \end{aligned} \tag{50}$$

Since (50) holds for all $i = 1, 2, \ldots, k$, we obtain, for all $||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2 < 2/K$

$$||\nabla(M(\boldsymbol{\theta}))||_2 \leq ||\frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}||_2 M(\boldsymbol{\theta}) \frac{2K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2}{2 - K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2} \leq ||\frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}||_2 M(\boldsymbol{\theta}) \frac{2K||\boldsymbol{\theta}||_2}{\sigma_{(1)}(2 - K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2)}. \tag{51}$$

Continuing as in the proof of Theorem 2.1, (51) yields that for all $\boldsymbol{\theta} \in \mathbb{R}^k$ with $||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2 < 2/K$, we have

$$||\nabla(\log(M(\boldsymbol{\theta})))||_2 = \frac{||\nabla M(\boldsymbol{\theta})||_2}{M(\boldsymbol{\theta})} \leq ||\frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}||_2 \frac{2K||\boldsymbol{\theta}||_2}{\sigma_{(1)}(2 - K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2)}.$$

Using the mean value theorem, for all $0 \preceq \boldsymbol{\theta} \in \mathbb{R}^k$ with $||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2 < 2/K$,

$$\log(M(\boldsymbol{\theta})) = \nabla(\log(M(\mathbf{z}))).\boldsymbol{\theta},$$

for some $\mathbf{0} \preceq \mathbf{z} \preceq \boldsymbol{\theta}$. Hence $||\mathbf{z}/\boldsymbol{\sigma}||_2 \leq ||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2 < 2/K$,

$$|\log(M(\boldsymbol{\theta}))| \leq ||\nabla(\log(M(\mathbf{z})))||_2 ||\boldsymbol{\theta}||_2 \leq ||\frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}||_2 \frac{2K||\mathbf{z}||_2}{\sigma_{(1)}(2 - K||\mathbf{z}/\boldsymbol{\sigma}||_2)} ||\boldsymbol{\theta}||_2. \tag{52}$$

Note that

$$||\boldsymbol{\theta}||_2 < \frac{1}{K_2} \Rightarrow ||\frac{\boldsymbol{\theta}}{\boldsymbol{\sigma}}||_2 < \frac{2}{K}.$$

Since $\mathbf{0} \preceq \mathbf{z} \preceq \boldsymbol{\theta}$, if $||\boldsymbol{\theta}||_2 < 1/K_2$, (52) yields

$$|\log(M(\boldsymbol{\theta}))| \leq ||\frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}}||_2 \frac{2K||\boldsymbol{\theta}||_2^2}{\sigma_{(1)}(2 - K||\boldsymbol{\theta}/\boldsymbol{\sigma}||_2)} \leq \frac{K_1||\boldsymbol{\theta}||_2^2}{2(1 - K_2||\boldsymbol{\theta}||_2)}.$$

Hence if $\boldsymbol{\theta} \succeq \mathbf{0}$ and $||\boldsymbol{\theta}||_2 < 1/K_2$, we obtain

$$P\left(\frac{\mathbf{W} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \succeq \mathbf{t}\right) \leq P\left(\boldsymbol{\theta}^t \frac{\mathbf{W} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \geq \boldsymbol{\theta}^t \mathbf{t}\right) \leq e^{-\boldsymbol{\theta}^t \mathbf{t}} M(\boldsymbol{\theta}) \leq \exp\left(-\boldsymbol{\theta}^t \mathbf{t} + \frac{K_1||\boldsymbol{\theta}||_2^2}{2(1 - K_2||\boldsymbol{\theta}||_2)}\right). \tag{53}$$

Using $\boldsymbol{\theta} = \mathbf{t}/(K_1 + K_2||\mathbf{t}||_2)$ in (53), we obtain

$$P\left(\frac{\mathbf{W} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \succeq \mathbf{t}\right) \leq \exp\left(-\frac{||\mathbf{t}||_2^2}{2(K_1 + K_2||\mathbf{t}||_2)}\right).$$

□

6 An application

Let $\tau_1, \tau_2 \in S_m$ be two fixed permutations from S_m, the permutation group on m elements. Let π be a permutation selected uniformly at random from S_n, where $n \geq m$. We consider the bivariate random variable $\mathbf{W} = (W_1, W_2)$ where W_1 counts the number of times pattern τ_1 appears in π and W_2 counts the number of times τ_2 appears in π. Concentration of measure inequalities for W_1 has been obtained in [7]. Using Theorem 5.1, we can in fact obtain concentration bounds for (W_1, W_2).

To fix notations, for $n \geq m \geq 3$, let π and τ be permutations of $\mathcal{V} = \{1, \ldots, n\}$ and $\{1, \ldots, m\}$, respectively, and let

$$\mathcal{V}_\alpha = \{\alpha, \alpha + 1, \ldots, \alpha + m - 1\} \quad \text{for } \alpha \in \mathcal{V},$$

where addition of elements of $\mathcal{V}$ is modulo n. We say the pattern τ appears at location $\alpha \in \mathcal{V}$ if the values $\{\pi(v)\}_{v \in \mathcal{V}_\alpha}$ and $\{\tau(v)\}_{v \in \mathcal{V}_1}$ are in the same relative order. Equivalently, the pattern τ appears at α if and only if $\pi(\tau^{-1}(v) + \alpha - 1), v \in \mathcal{V}_1$ is an increasing sequence. When $\tau = \iota_m$, the identity permutation of length m, we say that π has a rising sequence of length m at position α. Rising sequences are studied in [2] in connection with card tricks and card shuffling.

Letting π be chosen uniformly from all permutations of $\{1, \ldots, n\}$, and $X_{\alpha,\tau}$ the indicator that τ appears at α,

$$X_{\alpha,\tau}(\pi(v), v \in \mathcal{V}_\alpha) = 1(\pi(\tau^{-1}(1) + \alpha - 1) < \cdots < \pi(\tau^{-1}(m) + \alpha - 1)),$$

the sum $W = \sum_{\alpha \in \mathcal{V}} X_{\alpha,\tau}$ counts the number of m-element-long segments of π that have the same relative order as τ.

Let σ_α be the permutation of $\{1, \ldots, m\}$ for which

$$\pi(\sigma_\alpha(1) + \alpha - 1) < \cdots < \pi(\sigma_\alpha(m) + \alpha - 1).$$

$$\pi_1^\alpha(v) = \begin{cases} \pi(\sigma_\alpha(\tau_1(v - \alpha + 1)) + \alpha - 1), & v \in \mathcal{V}_\alpha \\ \pi(v) & v \notin \mathcal{V}_\alpha. \end{cases}$$

In other words π_1^α is the permutation π with the values $\pi(v), v \in \mathcal{V}_\alpha$ reordered so that $\pi_1^\alpha(\gamma)$ for $\gamma \in \mathcal{V}_\alpha$ are in the same relative order as τ_1. Similarly we can define π_2^α corresponding to τ_2.

To obtain $\mathbf{W}^i$, the $\mathbf{W}$ size biased variate in direction i for $i = 1, 2$, pick an index β uniformly from $\{1, 2, \ldots, n\}$ and set $W_j^i = \sum_{\alpha \in \mathcal{V}} X_{\alpha,\tau_j}(\pi_i^\beta)$. Then $\mathbf{W}^i = (W_1^i, W_2^i)$, for $i = 1, 2$.

The fact that we indeed obtain the desired size bias variates follows from results in [8]. Since both π_1^β and π_2^β agree with π on all the indices leaving out $\mathcal{V}_\beta$ and $|\mathcal{V}_\beta| = m$, we obtain $|W_j^i - W_j| \leq 2m - 1$ for $i, j = 1, 2$. Hence, $||\mathbf{W} - \mathbf{W}^i||_2 \leq (2m - 1)\sqrt{2}$ for $i = 1, 2$.

For $\tau \in S_m$, let $I_k(\tau)$ be the indicator that $\tau(1), \ldots, \tau(m - k)$ and $\tau(k + 1), \ldots, \tau(m)$ are in the same relative order. Following the calculations in [7], we obtain

$$\mu_i = E(W_i) = \frac{n}{m!} \quad \text{and} \quad \sigma_i^2 = \text{var}(W_i) = n\left(\frac{1}{m!}\left(1 - \frac{2m - 1}{m!}\right) + 2\sum_{k=1}^{m-1} \frac{I_k(\tau_i)}{(m + k)!}\right).$$

Since $0 \leq I_k \leq 1$, the variance lower bound is obtained when $I_k = 0$ yielding

$$\sigma_{(1)}^2 \geq \frac{n}{m!}\left(1 - \frac{2m - 1}{m!}\right).$$

Since, the constants K_1 and K_2 Theorem 5.1 can be replaced by larger constants, we can apply it with

$$K_1 = \frac{(8m - 4)m!}{m! - 2m + 1} \quad \text{and} \quad K_2 = \frac{(2m - 1)m!}{\sqrt{2n(m! - 2m + 1)}},$$

to obtain concentration inequality for $\mathbf{W} = (W_1, W_2)$.

References

[1] Adamaczak, R. (2006), Moment inequalities for U statistics, *Ann. Probab.*, **34**, 2288-2314.

[2] Bayer, D. and Diaconis, P.(1992). Trailing the Dovetail Shuffle to its Lair. *Ann. of Appl. Probab.* **2**, 294-313.

[3] Chatterjee, S. (2007). Stein's method for concentration inequalities, *Probab. Theory Related Fields*, **138**, 305-321.

[4] Friedman, J. and Rafsky, L.C. (1979). Multivariate generalisations of the Wald-Wolfowitz and Smirnov two sample tests, *Ann. Statist.*, **7**, 697-717.

[5] Friedman, J. and Rafsky, L.C. (1983). Graph-theoretic measures of multivariate association and prediction, *Ann. Statist.*, **11**, 377-391.

[6] Ghosh, S. and Goldstein, L. (2009). Concentration of measures via size biased couplings, to appear in *Probab. Th. Rel. Fields.*

[7] Ghosh, S. and Goldstein, L.(2010). Applications of size biased couplings for concentration of measures, *preprint*.

[8] Goldstein, L.(2005). Berry Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing, *Journal of Applied Probability*, **42**, 661-683.

[9] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Ann. Math. Statist.*, **19**, 293-325.

[10] Houdré, C. and Reynaud-Bouret, P. (2003). Exponential inequalities with constants, for U statistics of order two. In *Stochastic Inequalities and Applications*, 55-69. Progr. Prob. 56. Birkhäuser, Basel.

[11] Lee, A.J. (1990). *U- statistics: Theory and practice*, Dekker, New York.

[12] Mann, H.B. and Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other, *Ann. Math. Statist.*, **18**, 50-60.

[13] Raič, M. (2007). CLT related large deviation bounds based on Stein's method, *Adv. Appl. Prob.*, **39**, 731-752.

[14] Reinert, G. and Röllin, A. (2008). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition, *Ann. Probab.*, **37**, 2150-2173.

[15] Reinert, G. and Röllin, A. (2009). U-statistics and random subgraph counts: Multivariate normal approximation via exchangeable pairs and embedding, *preprint*.

[16] Rinott,Y. and Rotar, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications, *Ann. Appl. Probab.*, **7**, 1080-1105.

[17] Stein, C. (1986). *Approximate computation of expectations*, Institute of Mathematical Statistics, Hayward, CA.

[18] Zhao, L., Bai, Z., Chao, C. and Liang, W. (1997). Error bound in a central limit theorem of double-indexed permutation statistics, *Ann. Statist.*, **25**, 2210-2227.